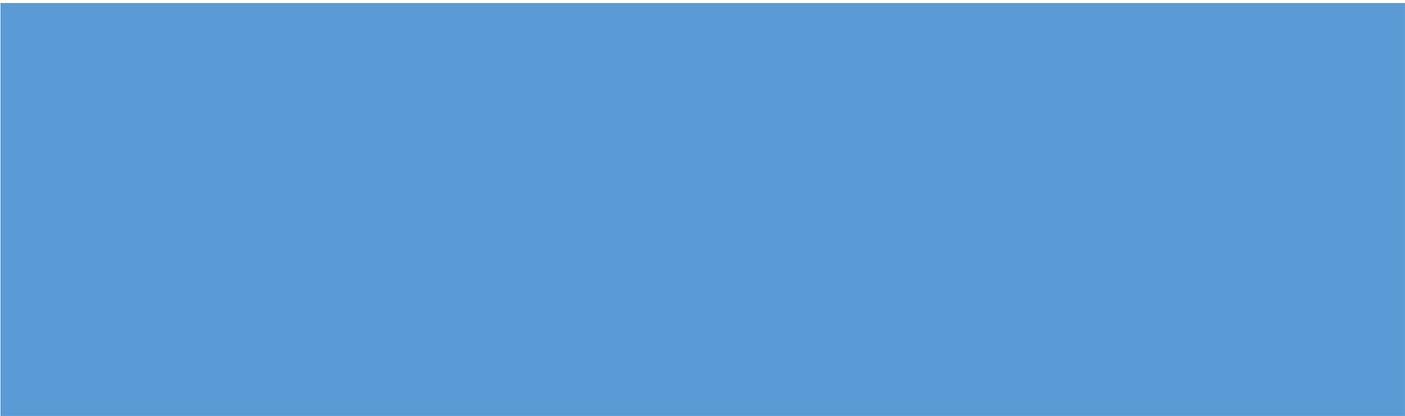




# Plan de Gestión de Datos



Proyecto PID2020-116540RB-C22

financiado por MCIN/ AEI /10.13039/501100011033

## VERSIÓN DEL DOCUMENTO

Fecha	Versión	Editor	Comments	Status
28/2/2022	v1	C. Guardiola Albert		Borrador
15/3/2022	V2	C. Guardiola Albert	Basado en los comentarios de M. Gómez, M. Sanabria, M. Béjar, O. Monserrat y M. Crossetto	Definitivo
31/8/2023	V3	C. Guardiola Albert	Actualización sobre los repositorios digitales del CSIC	Definitivo

## Índice de Contenidos

<b>VERSIÓN DEL DOCUMENTO</b> .....	1
<b>Plan de Gestión de Datos</b> .....	3
Detalles del proyecto [Project Details] .....	3
1. Resumen de datos [Data summary].....	4
<b>1.1. Propósito de la recopilación y generación de datos</b> [What is the purpose of the data collection/generation?] .....	4
<b>1.2. Relación de los datos con los objetivos del proyecto</b> [What is the relation of the data to the objectives of the project?] .....	4
<b>1.3. Formato y origen de datos generados / recopilados</b> [What types and formats of data will the project generate/collect? What is the origin of the data?] .....	5
<b>1.4. Reutilización de los datos</b> [Will you re-use any existing data and how?] .....	8
<b>1.5. Tamaño esperado de los datos</b> [What is the expected size of the data?] .....	8
<b>1.6. Impacto y utilidad de los datos</b> [To whom might it be useful (data utility)?].....	8
2. Datos FAIR [FAIR data] .....	11
<b>2.1. Capacidad de los datos de ser localizados [Findable Data]</b> .....	11
<b>2.2. Accesibilidad de los datos [Accesible Data]</b> .....	12
<b>2.3. Datos Interoperables [Interoperable data]:</b> .....	13
<b>2.4. Reutilización de los datos [Re-usable data]:</b> .....	14
3. Asignación de recursos [Allocation of resources].....	16
<b>3.1. Recursos humanos y económicos para garantizar que los datos de investigación cumplan con los principios FAIR</b> [What are the costs for making data FAIR in your project?] .....	16
<b>3.2. Responsables de la gestión de datos en el proyecto</b> [Who will be responsible for data management in your project?] .....	16

4.	Almacenamiento y salvaguardia de seguridad [Data security].....	17
	<b>4.1. Recuperación datos, almacenamiento seguro</b> [What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?]	17
5.	Aspectos éticos [Ethical aspects] .....	17
	<b>5.1. Aspectos éticos y protección de datos</b> [Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA). Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?]	17
6.	Otros [Others] .....	18
	<b>6.1. Procedimientos, políticas o referencias</b> .....	18

## Plan de Gestión de Datos

El presente Plan de Gestión de Datos (PGD) describe el ciclo de vida de la gestión de todos aquellos conjuntos de datos que sean recopilados, procesados o generados por el proyecto de investigación “Hacia una explotación inteligente de los datos de movimientos del terreno para la prevención y mitigación de riesgos geológico-geotécnicos (SARAI)”. En este documento se describe cómo serán almacenados y preservados los datos en el transcurso del proyecto de investigación e incluso tras su conclusión, describiendo qué datos serán recopilados, procesados o generados, qué metodología y estándares se utilizarán, cómo se compartirán y serán abiertos, y por último cómo se conservarán y preservarán. El PGD es un documento vivo, que irá evolucionando y adquiriendo más precisión y entidad durante el período de vigencia del proyecto. A falta de una plantilla de PGD dentro del Plan Nacional se ha intentado que el PGD se ajuste a la plantilla proporcionada por la Comisión Europea para la gestión de proyectos Horizon 2020. El PGD se actualizará al menos una vez hacia la mitad del transcurso del proyecto y otra al final para hacer los ajustes necesarios sobre los datos generados y los usos identificados por los responsables del proyecto, ya que no todos los datos o los usos potenciales que se pueda hacer de ellos están claros desde el principio. Se deberán crear nuevas versiones del PGD siempre que haya cambios importantes en el proyecto debidos a la inclusión de nuevos conjuntos de datos o a factores externos.

### Detalles del proyecto [Project Details]

- **Nombre del proyecto:** Hacia una explotación inteligente de los datos de movimientos del terreno para la prevención y mitigación de riesgos geológico-geotécnicos (SARAI)
- **Resumen del proyecto:** Este proyecto tiene como objetivo mejorar las capacidades de explotación de las técnicas de teledetección basadas en radar para medir y monitorear la deformación de la tierra: SAR interferométrico diferencial (DInSAR) e interferometría de dispersión persistente (PSI). El lanzamiento de las misiones Sentinel-1A y 1B del Programa Copernicus marcó un cambio de paradigma: los datos SAR estarán disponibles durante muchos años a partir de ahora en todo el mundo y de forma gratuita. En 2016, una iniciativa europea planteó implementar un Servicio Europeo de Movimiento de Tierra (EGMS) basado en DInSAR para proporcionar datos consistentes a escala continental. El proyecto SARAI quiere mejorar la explotación de los próximos resultados del EGMS, mediante el desarrollo de herramientas y procedimientos para procesar y analizar los principales productos de DInSAR y PSI (i.e. la velocidad de deformación lineal durante el período observado y las series de tiempo). Para facilitar el uso de estos datos a usuarios no expertos, SARAI creará herramientas para: (i) Identificar las señales de deformación y separarlas del ruido; (ii) Clasificar estas señales según el fenómeno causante, natural o antrópico; (iii) Estimar las áreas más susceptibles de experimentar movimientos del terreno; (iv) Predecir la evolución temporal de la deformación del terreno
- **Fecha de principio y fin del proyecto:** 1/9/2021 – 31/8/2025
- **ID:** PID2020-116540RB-C22
- **Financiador del proyecto:** Ministerio de Ciencia e Innovación
- **URL:** <http://maps.igme.es/sarai/>

### Contribuidores [Contributors]

Los investigadores principales del proyecto y responsables de la correcta gestión de los datos, son por parte del Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) **Oriol Monserrat Hernández** ([oriol.monserrat@cttc.cat](mailto:oriol.monserrat@cttc.cat)) y **Michele Crosetto** ([michele.crosetto@cttc.cat](mailto:michele.crosetto@cttc.cat)), y por parte del Centro Nacional del CSIC Instituto Geológico y

## 1. Resumen de datos [Data summary]

### 1.1. Propósito de la recopilación y generación de datos [What is the purpose of the data collection/generation?]

El propósito, tanto de la recopilación de datos que se va a llevar a cabo en el presente proyecto, como de la generación de nuevos datos, no es otro que el que persigue el propio proyecto SARAI: mejorar las capacidades de explotación de los datos de la deformación de la tierra medidos con la técnica SAR interferométrico diferencial (DInSAR) e interferometría de dispersión persistente (PSI).

### 1.2. Relación de los datos con los objetivos del proyecto [What is the relation of the data to the objectives of the project?]

A continuación, se explica la relación de los objetivos específicos del proyecto SARAI con los datos recopilados y generados durante el proyecto, gestionados según el presente PGD:

- O1. Desarrollo de técnicas DInSAR y PSI (responsable O. Monserrat). El objetivo es mejorar la cadena existente del CTTC para el procesamiento de este tipo de datos. Los datos recopilados serán los propios datos preprocesados DInSAR y PSI, que a través de diferentes técnicas desarrolladas por el CTTC lograrán disminuir los errores del procesado DInSAR y PSI resultante (datos generados).
- O2. Explotación de los resultados de DInSAR/PSI (responsable O. Monserrat). Existe la necesidad de mejorar herramientas automáticas de análisis de datos para explotar el conjunto de datos DInSAR y PSI mencionados en el objetivo O1.
- O3. Explotación de datos EGMS (responsable M. Crosetto). El EGMS proporcionará una cobertura europea de productos derivados de DInSAR. Estos datos no son generados ni recopilados por el proyecto SARAI, sino que serán utilizados en base a consultas que se realicen del EGMS, para desarrollar herramientas que mejoren la interpretación de estos datos.
- O4. Generación de una base de datos de deformación InSAR (responsable M. Béjar). Se creará una base de datos que contenga puntos de medición de movimiento del terreno clasificados (PS) con la deformación asociada (velocidad media y series temporales) y un conjunto de covariables ambientales, con dos objetivos principales: 1) Crear un inventario de deformación InSAR clasificada basada en todos los resultados de procesamiento obtenidos por los equipos de IGME y CTTC en proyectos anteriores; 2) La base de datos se utilizará como conjunto de datos de entrenamiento para desarrollar los algoritmos de inteligencia artificial para el O5. El número y tipo de covariables evolucionará a medida que avance el proyecto.
- O5. Desarrollo de herramientas basadas en inteligencia artificial para clasificar la deformación activa (responsable C. Guardiola). Se utilizarán diferentes herramientas de aprendizaje automático que se entrenarán con la base de datos creada según el O4. Los PS clasificados según el algoritmo desarrollado serán datos nuevos generados en el proyecto, que serán almacenados como mapas, polígonos o puntos clasificados en función de la eficiencia obtenida.

- 06. Desarrollo de herramientas basadas en inteligencia artificial para generar mapas de susceptibilidad al movimiento del suelo (responsable M. Béjar) Las probabilidades de resultantes del clasificador de aprendizaje automático obtenido en el objetivo O5, así como y los pesos las covariables estudiadas, se utilizarán para desarrollar índices de susceptibilidad para diferentes riesgos geológicos a escala nacional. Los resultados serán mapas de susceptibilidad de diferentes riesgos geológicos a escala nacional.
- 07. Predicción de la deformación del terreno utilizando inteligencia artificial (C. Guardiola). En determinadas zonas piloto se generarán series temporales futuras de deformación con una incertidumbre asociada a la predicción.

**1.3. Formato y origen de datos generados / recopilados** [What types and formats of data will the project generate/collect? What is the origin of the data?]

En la Tabla 1 se indican los formatos de los datos, donde se ha procurado usar siempre formatos estándares, ya que asegura la usabilidad de los datos a largo plazo.

Según se ha descrito en el punto anterior, el objetivo O4 del proyecto SARAI es la construcción de una base de datos, para lo que se recopilará una cantidad importante de datos en la Tabla 1. La propia evolución de las actividades del proyecto supondrá una modificación en esta Tabla 1, ya que habrá datos que podrían descartarse si tuvieran poca relación con las medidas de los movimientos del terreno, o su imposibilidad **de usarlas. Igualmente, podrán existir datos no considerados en esta fase inicial del proyecto, que más adelante se demuestren sean necesarios. Por todo ello la Tabla 1 deberá revisarse en las próximas actualizaciones del PDG.**

**Tabla1. Datos utilizados (recopilados y generados) en el proyecto SARAI.**

Datos	Tipo/Formato	Origen
Pre process InSAR/PSI data	Satellite images: S1 images in SAFE format (image data in various binary formats, 'quicklooks' in PNG format, Google Earth overlays in KML format and HTML preview files, product metadata and other information in XML) DEMs (ArcInfo ASCII and GeoTiff format), orbits (ascii)	CTTC
IGME and CTTC InSAR results: <i>ID PS</i> <i>PS lon, lat</i> <i>PS mean_VLOS</i> <i>PS time series (dates and displacement)</i> <i>PS time series type (linear, etc)</i> <i>PS time series metadata (ini, end, #)</i> <i>PS deformation range</i> <i>Error, std, etc</i> <i>Spatial resolution</i>	Vector (shape, csv, ascii) / raster, Data base (Timescale / hyper table)	IGME's and CTTC's InSAR published results

Datos	Tipo/Formato	Origen
<i>Metadata (asc/desc, sat, incidence, azimuth, software, track)</i> Interferograms		
Sentinel-1 SAR data	Satellite images, vector (shape, csv)	<a href="https://scihub.copernicus.eu/dhus/">https://scihub.copernicus.eu/dhus/</a>
Paz SAR data	Satellite images, vector (shape, csv)	Paz data proposal (AO-0001-026) - IGME User License for Scientific Use
EGMS results	Vector (shape, csv, ascii), raster	<a href="https://land.copernicus.eu/pan-european/european-ground-motion-service">https://land.copernicus.eu/pan-european/european-ground-motion-service</a>
GEP/EPOS InSAR results	Vector (shape, csv)	<a href="https://geohazards-tep.eu/#!">https://geohazards-tep.eu/#!</a>
Geology/Lithology	Vector (shape)	<a href="https://info.igme.es/cartografiadigital/geologica/Geode.aspx?languaje=es">https://info.igme.es/cartografiadigital/geologica/Geode.aspx?languaje=es</a>
Active faults nearby	Vector (shape)	<a href="http://info.igme.es/qafil/">http://info.igme.es/qafil/</a>
MDE	ASCII matriz ESRI (.asc)	<a href="http://centrodedescargas.cnig.es/CentroDescargas/index.jsp">http://centrodedescargas.cnig.es/CentroDescargas/index.jsp</a>
Slope	ASCII ESRI (.asc)	<a href="http://centrodedescargas.cnig.es/CentroDescargas/index.jsp">http://centrodedescargas.cnig.es/CentroDescargas/index.jsp</a>
Land cover	Vector (shape)	Corine CLC 2018 <a href="https://land.copernicus.eu/pan-european/corine-land-cover/clc2018">https://land.copernicus.eu/pan-european/corine-land-cover/clc2018</a>
Aquifer beneath the PS (Masa de Agua Subterránea)	Vector (shape)	<a href="https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/agua/masas-de-agua-phc-2015-2021.aspx">https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/agua/masas-de-agua-phc-2015-2021.aspx</a>
IGME groundwater database	Data base (Access)	Base datos de aguas del IGME

Datos	Tipo/Formato	Origen
		<a href="https://info.igme.es/BDAguas/">https://info.igme.es/BDAguas/</a>
MITECO groundwater database	Data base (Access)	Red de control del estado cuantitativo de las aguas subterráneas del MITECO <a href="https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/agua/red-piezometrica.aspx">https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/agua/red-piezometrica.aspx</a>
Spanish Land Movements database	Data base (Access)	Base de datos MOVES del IGME <a href="https://info.igme.es/BD2DMoves/">https://info.igme.es/BD2DMoves/</a>
Precipitation	Ascii (.txt)	Servicio de descargas AEMET <a href="http://www.aemet.es/es/serviciosclimaticos/ambito_climat/datos_diarios">http://www.aemet.es/es/serviciosclimaticos/ambito_climat/datos_diarios</a>
Spanish mining cadastre	Data base (Access)	<a href="https://geoportal.mine.tur.gob.es/CatastroMinero/BusquedaAlfanumerica">https://geoportal.mine.tur.gob.es/CatastroMinero/BusquedaAlfanumerica</a>
Industrial areas, Solid Urban Waste	Cartography (vector/shape)	Regional IDEE
Transport Network	Cartography (vector/shape)	<a href="http://centrodedescargas.cnig.es/CentroDescargas/index.jsp">http://centrodedescargas.cnig.es/CentroDescargas/index.jsp</a>
New processed InSAR/PSI data	Vector (shape, csv, ascii), raster	New SARAI product
Classification of PS	Data base (Timescale / hyper table)	New SARAI product
Land movements susceptibility maps	Cartography (vector/shape)	New SARAI product
Predicted land movements	Spatio temporal information (.txt/.csv)	New SARAI product
Scientific papers	.pdf	New SARAI product

#### 1.4. Reutilización de los datos [Will you re-use any existing data and how?]

Como se ha mencionado en los apartados anteriores la mayoría de los datos que se van a utilizar durante el presente proyecto provienen de fuentes externas al proyecto SARAI (ver origen de los datos en la Tabla 1). Los datos utilizados corresponden todos a datos públicos o datos resultantes de proyectos anteriores de los participantes en el proyecto SARAI.

La reutilización de todos estos datos dará lugar a nuevos productos. De forma simplificada estos nuevos productos están reflejados en las últimas filas de la Tabla 1, se listan a continuación:

1. Procesados mejorados de los datos InSAR/PSI [New processed InSAR/PSI data] (formato vector/shape, csv, ascii, raster).
2. Base de datos con dimensión temporal que recogerá para cada punto PS la mayoría de la información recopilada en la Tabla 1 (formato de base de datos “anidada” con series temporales).
3. Clasificación de los PS realizada de forma automática (formato de base de datos “anidada” con series temporales) [Classification of PS].
4. Mapas de susceptibilidad de movimientos del terreno (formato vector/shape) [Land movements susceptibility maps].
5. Predicciones de movimientos del terreno en zonas piloto (formato .txt/.csv) [Predicted land movements].
6. Artículos científicos (pdf) [Scientific papers].

Los datos descritos en los puntos del 2 al 5 se realizarán durante el proyecto SARAI y serán publicados al finalizar el mismo. La base de datos a la que se refiere el punto 1 tiene vocación de ser una base de datos que se mantenga viva cuando finalice el proyecto.

#### 1.5. Tamaño esperado de los datos [What is the expected size of the data?]

En la actualidad estamos evaluando de forma aproximada el volumen que van a ocupar los datos en MB / GB / TB. Se esperan tener dos tipos de volúmenes de datos: los derivados de los datos PS que implican un gran volumen que puede ir del orden de GB a TB, y que se está estudiando como almacenarlos para uso eficiente del producto, y el resto de datos que no superaran los 2GB. En las siguientes actualizaciones del PDG podremos dar detalles sobre las proporciones de datos brutos y datos procesados de los datos que aparecen en la Tabla 1, así como de otras salidas secundarias como pueden ser los artículos científicos. También se podrán incluir en futuras actualizaciones las implicaciones de los volúmenes de datos en términos de almacenamiento, acceso y conservación, por si fuera necesario incluir costes adicionales

#### 1.6. Impacto y utilidad de los datos [To whom might it be useful (data utility)?]

Los productos del proyecto SARAI detectarán áreas susceptibles a diferentes peligros geológicos relacionados con la deformación del suelo. En consecuencia, los resultados de SARAI apoyarán las políticas nacionales, regionales y locales involucradas en el mapeo y la gestión de riesgos, lo que permitirá una mejor preparación para el cambio climático y una provisión más sólida para la protección de la sociedad y la seguridad de las personas contra los riesgos geológicos. Otras aplicaciones finales son la geodesia y

la gestión del territorio, la planificación urbana y rural, la protección del patrimonio cultural, los servicios climáticos, el desarrollo de infraestructura, la extracción de minería y otros recursos naturales, el monitoreo de presas y aguas subterráneas, los temas de seguros y litigios, la ingeniería estructural y civil, el mercado inmobiliario y el transporte. Dentro de los usuarios finales se incluyen estudios geológicos, administraciones de carreteras, ferrocarriles y minería, reguladores y planificadores, autoridades a nivel europeo, nacional, regional y municipal, ciudadanos de los estados participantes de Copernicus, PYMEs e industria. Los impactos específicos esperados y cómo se lograrán se destacan en la Tabla 2. La innovación que aportan los productos del proyecto SARAI se dirige a múltiples beneficiarios de la ciencia, el medio ambiente, la sociedad, el progreso tecnológico y la economía/competitividad (Tabla 3).

**Tabla 2. Impactos esperados (EI, por sus siglas en inglés) de los productos del proyecto SARAI y su relación con los objetivos del proyecto (O).**

Impacto esperado	¿Cómo lograrán estos EI los productos de SARAI?
Mejorar las capacidades para evaluar y mitigar los riesgos geológicos en áreas urbanas, infraestructuras y ecosistemas (EI1).	<ul style="list-style-type: none"> <li>- Datos de desplazamiento actualizados con herramientas mejoradas de procesamiento y análisis de DInSAR y PSI: separando las señales de deformación y el ruido. (O1)</li> <li>- Herramientas de análisis automático para explotar los datos DInSAR y su relación con los componentes básicos de la variabilidad espaciotemporal del movimiento del suelo. (O2, O3, O5)</li> <li>- Nuevo inventario de deformación clasificada de InSAR basada en todos los resultados de procesamiento obtenidos por los equipos de IGME y CTTC en proyectos anteriores (O4)</li> </ul>
Mejorar la predicción de riesgos geológicos en zonas urbanas e infraestructuras (EI2)	<ul style="list-style-type: none"> <li>- Herramientas para la predicción y detección temprana de deformaciones/riesgos geológicos del suelo en áreas específicas utilizando inteligencia artificial (O7)</li> <li>- Mapas de susceptibilidad basados en la probabilidad para diferentes riesgos geográficos a escala nacional (O6)</li> </ul>
Mejorar las estrategias de gestión de infraestructuras civiles para apoyar a los responsables de la toma de decisiones (EI3) Mejorar la resiliencia y la seguridad para el peligro geológico relacionado con la deformación de la tierra (EI4)	<ul style="list-style-type: none"> <li>- Mapas de deformación de la tierra que puedan mostrar áreas de riesgo geológico potencial. Clasificación de estas señales según el fenómeno causal natural o antrópico (O5)</li> <li>- Estimar las áreas más susceptibles a experimentar movimientos de tierra en el futuro (O6)</li> <li>- Evolución temporal futura a corto plazo de la deformación del suelo (O7)</li> </ul>
Nuevas rutinas de modelado para determinar y predecir automáticamente el movimiento del suelo a partir de datos de InSAR	<ul style="list-style-type: none"> <li>- Herramientas basadas en inteligencia artificial para clasificar la deformación activa (O5)</li> <li>- Herramientas basadas en inteligencia artificial para la predicción del movimiento del suelo (O7)</li> </ul>

(E15)	
-------	--

**Tabla 3. Beneficiarios previstos en Ciencia, Medio Ambiente, Sociedad, Progreso tecnológico, Economía/competitividad. Entidades promotoras observadoras (EPO) que apoyan el proyecto SARAI.**

Tipo	Sector	EPO	Oportunidades que dan los productos de SARAI y que están más allá del impacto directo
Ciencia	Universidad, instituciones de investigación (pública) tanto a nivel nacional como internacional	- Uc3m-Santander Big Data Institute	Contribuir al conocimiento científico de la caracterización de los movimientos terrestres a partir de observaciones satelitales.
Sociedad	Planificación del terreno (pública)	- Instituto Geográfico Nacional	Proporcionar información sobre la deformación del suelo que mejore la planificación del uso de la tierra y disminuya el coste para los municipios.
Sociedad	Gestión de los riesgos geológicos (pública)	- Earth Observation and Geohazards Expert Group from the Geological Surveys of Europe	Mejorar los planes dedicados a la protección contra los peligros naturales (por ejemplo, deslizamientos de tierra) y antropogénicos (por ejemplo, el impacto de los trabajos de excavación).
Medio ambiente	Gestión de los riesgos por desastres medioambientales (pública)	- Ministerio para la Transición Ecológica y Reto Demográfico - WWF Adena	Localizar ecosistemas en riesgo debido a la sobreexplotación del agua (que desencadena el hundimiento de la tierra) o los movimientos del suelo.
Progreso tecnológico y económico	Compañías especializadas en <i>remote sensing</i> & <i>InSAR</i> companies (privada)	- Geokinesia (star-up) - Detektia (start-up) - Dares	Mejorar las herramientas que se utilizan para analizar los datos de InSAR.
Progreso tecnológico y económico	Compañías de tecnología digital (privada)	- Singular People S.L.	Promover el desarrollo de empresas de tecnología digital (Big data, IA, Agile, Blockchain, etc.).

Tipo	Sector	EPO	Oportunidades que dan los productos de SARAI y que están más allá del impacto directo
Progreso tecnológico y económico	Responsables de infraestructuras (público/privada)	- COMSA	Motivar a las entidades encargadas de las infraestructuras civiles a utilizar la tecnología basada en InSAR como herramienta de control.
Progreso tecnológico y económico	Consortios y compañías de seguros (público/privada)	- Consorcio de Compensación de Seguros	Activar el uso de la tecnología InSAR para compañías de seguros (sector privado).

## 2. Datos FAIR [FAIR data]

Los datos generados en el proyecto SARAI siguen los principios **FAIR (Findable, Accesible, Interoperable, Reusable)**, que son un conjunto de principios rectores para hacer que los datos de investigación sean fáciles de encontrar, accesibles, interoperables y reutilizables (Wilkinson et al., 2016).

### 2.1. Capacidad de los datos de ser localizados [Findable Data]

**2.1.1. Capacidad de descubrimiento de datos generados por SARAI (provisión de metadatos y los estándares para la creación de estos)** [Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)? What metadata will be created?]

Los metadatos se elaborarán intentando seguir el formato INSPIRE TG2.0 en la medida en que los datos lo permitan, asegurando así el descubrimiento de los datos (en inglés *data discovery*). Se aportarán metadatos para los siguientes productos listados en el apartado 1.4:

- *Base de datos con dimensión temporal que recogerá para cada punto PS la mayoría de la información recopilada en la Tabla 1:* Descripción de cada una de las variables incluidas en dicha base de datos, formato y origen de los datos recopilados. Habrá que estudiar si esta base de datos puede formar parte de la base institucional del IGME-CSIC.
- *Clasificación de los PS realizada de forma automática (formato de base de datos "anidada" con series temporales):* Descripción de las categorías incluidas en la clasificación y referencia de la metodología.
- *Mapas de susceptibilidad de movimientos del terreno:* descripción de los valores de susceptibilidad y referencia de la metodología utilizada.
- *Predicciones de movimientos del terreno en zonas piloto:* descripción de las variables incluidas en la serie temporal y su formato.

Los datos albergados en el repositorio Zenodo incluirán un fichero *readme* junto con los diferentes tipos de *dataset*, donde se incluirán los metadatos.

**2.1.2. Identificadores persistentes y únicos como los identificadores de objetos digitales** [What naming conventions do you follow?]

El repositorio principal para el almacenamiento de los datos abiertos será ZENODO, por lo que se utilizarán los identificadores de objeto digital (DOI por sus siglas en inglés), que genera ZENODO tanto para artículos de revistas, conjuntos de datos o versiones de código abierto.

**2.1.3. Nomenclatura, palabras clave y control de versiones** [Will search keywords be provided that optimize possibilities for re-use? Do you provide clear version numbers?]

La nomenclatura utilizada y palabras clave para aplicar en estrategias de búsqueda serán desarrolladas en las futuras actualizaciones del presente PDG. El proyecto SARAI tiene una duración de 4 años. Al no ser un proyecto con continuidad, la versión de los datos tendrá una única versión, pudiendo este hecho verse modificado por posibles localizaciones de errores u erratas, o por la existencia de futuros proyectos que den continuidad a algunas de las tareas de este.

**2.2. Accesibilidad de los datos [Accessible Data]**

**2.2.1. Disponibilidad de los datos** [Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions]

Los productos de las investigaciones promovidas dentro del proyecto SARAI y relacionados con los objetivos O3, O4, O5, O6 y O7 tendrán un carácter abierto, siempre y cuando hayan sido validados y se hayan publicado los trabajos científicos relacionados con dichos datos y propuestos como hitos dentro del proyecto. Los datos producidos por el proyecto SARAI, que estén relacionados con los objetivos O1 y O2, estarán albergados en los servidores del CTTC, por el gran tamaño de los datos.

**2.2.2. Repositorios donde se van a depositar los datos** [How will the data be made accessible (e.g. by deposition in a repository)?]

Los datos producto del proyecto SARAI relacionados con los objetivos del O3 al O7 se almacenarán en el repositorio ZENODO (<https://zenodo.org/>), mientras que las publicaciones científicas se almacenarán en Digital CSIC (<https://digital.csic.es/>). Los datos producto de los objetivos O1 y O2 estarán almacenados en los servidores del CTTC y estarán podrán ser accesibles a petición justificada de los mismos.

Se intentará que la base de datos generada forme parte de las bases de datos institucionales del IGME. En ese caso, la base de datos residirá también en los servidores del CSIC y los metadatos en el catálogo de metadatos del IGME-CSIC. Se evaluará así mismo la viabilidad de publicar la base de datos y el mapa de susceptibilidad bajo un servicio de mapa.

**2.2.3. Métodos o herramientas de software necesarios para acceder a los datos** [What methods or software tools are needed to access the data?]

El software necesario para acceder a los datos será diferente según el formato que tengan los mismos:

1. *Base de datos con dimensión temporal que recogerá para cada punto PS la mayoría de la información recopilada en la Tabla 1:* en estudio, una opción que se está manejando es SQL

2. *Procesados mejorados de los datos InSAR/PSI*: se definirán en próximas versiones del PGD.
3. *Clasificación de los PS realizada de forma automática (formato de base de datos "anidada" con series temporales)*: SQL
4. *Mapas de susceptibilidad de movimientos del terreno*: Sistema de información geográfica tipo ArcGIS o Qgis
5. *Predicciones de movimientos del terreno en zonas piloto*: hoja de cálculo o editor de texto o SQL, formato a definir en función del tamaño de los datos que se generarán
6. *Publicaciones científicas*: lector de pdf

**2.2.4. Lugar de depósito de los datos y los metadatos asociados, la documentación y los códigos** [Where will the data and associated metadata, documentation and code be deposited?]

En los repositorios mencionados, en el CTTC y en el CSIC. Además, se estudiará la viabilidad de utilizar servidores del IGME-CSIC y el catálogo de metadatos del IGME-CSIC.

**2.2.5. Acceso a los datos con algún tipo restricción** [If there are restrictions on use, how will access be provided? Is there a need for a data access committee? Are there well described conditions for access (i.e. a machine readable license)? How will the identity of the person accessing the data be ascertained?]

El acceso a los datos no depositados en los repositorios se solicitará mediante correo electrónico a los responsables del proyecto, que evaluarán y aprobarán dicho acceso. Se establecerán las condiciones específicas de utilización de los datos y se firmará una carta de compromiso o acuerdo contractual.

**2.3. Datos Interoperables [Interoperable data]:**

**2.3.1. Datos y vocabularios de metadatos, estándares o metodologías se siguen para facilitar la interoperabilidad** [What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?]

El concepto de interoperabilidad se aplica tanto en el nivel de los datos como en el de los metadatos (véase formato de los metadatos). Se tendrán en cuenta las reglas de aplicación de INSPRIE sobre los metdatos, que se corresponden con las normas ISO sobre los metadatos. La interoperabilidad consiste en el intercambio y la reutilización de datos entre investigadores, instituciones, organizaciones, países, etc. (utilizar formatos estándares, y, siempre que sea posible, hacer el acceso a los datos a través de software abierto). Los datos deberían poder combinarse y usarse con otros datos o herramientas. Para ello el formato de los datos será abierto e interpretable por distintas herramientas, incluido el software abierto como Qgis, R o Python (véase apartado 2.2.3).

**2.3.2. Vocabulario estándar para todos los datos de dataset** [Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?]

En las próximas actualizaciones del PDG, cuando se haya avanzado en las actividades de investigación, se podrá evaluar la necesidad de definir un vocabulario estándar para todos los datos del proyecto SARAI, así como la creación de este si así fuese. En todo caso se intentará utilizar la semántica de los

conjuntos de datos utilizando las especificaciones de datos de INSPIRE o las acordadas internacionalmente.

## **2.4. Reutilización de los datos [Re-usable data]:**

### **2.4.1. Tipo de licencia que va a utilizar para los datos generados en el proyecto SARAI** [How will the data be licensed to permit the widest re-use possible?]

La licencia asignada se intentará que sea la menos restrictiva, pero **se definirá en próximas actualizaciones del PDG**. El tipo de licencias podrían ser:

- En Licencias Creative Commons, se utilizará CC BY (permitir uso comercial y permitir modificaciones).
- En Licencias Open Data Commons (licencias específicas para datos), se utilizará Public Domain Dedication and License (PDDL — “Public Domain for data/databases”), permite compartir libremente, modificar y usar el trabajo sin restricciones.
- Licencias para código informático o software.

Los datos albergados en las bases de datos institucionales o los servidores del IGME-CSIC estarán accesibles y reutilizables bajo la licencia de datos del IGME-CSIC: <https://www.igme.es/Ayuda/CondiUso.htm>.

### **2.4.2. Periodo de embargo a los datos**

Tal y como se ha mencionado en el apartado 2.2.1. los datos tendrán un carácter abierto una vez hayan sido validados y se hayan realizado las publicaciones científicas comprometidas en los diferentes hitos del proyecto. A excepción de los datos de los objetivos O1 y O2.

Las publicaciones científicas estarán sometidas a las condiciones de la licencia de cada revista. Algunas de las mismas tienen un periodo de embargo, que depende de la editorial y la revista. En todo caso, las versiones permitidas por estas licencias serán almacenadas en el repositorio mencionado en el punto 2.2.2.

### **2.4.3. Uso por terceras personas** [Are the data produced and/or used in the project useable by third parties, in particular after the end of the project?]

Una vez haya finalizado el proyecto SARAI se permite el uso por terceras personas de los datos creados o usados en el proyecto.

La utilización de datos e información del proyecto SARAI, se realizará por parte de los usuarios o agentes de la reutilización bajo su propia cuenta y riesgo, correspondiéndoles en exclusiva responder frente a terceros por daños que pudieran derivarse de ella.

El CTTC, el IGME-CSIC, y los responsables del proyecto SARAI no serán responsables de ninguna pérdida, reclamación o perjuicio derivado del uso o mal uso que de su información hagan los agentes reutilizadores ni tampoco de los daños sufridos o pérdidas económicas que, de forma directa o indirecta, produzcan o puedan producir perjuicios económicos, materiales o sobre datos, provocados por el uso de la información reutilizada o las decisiones derivadas de su utilización o manipulación.

El CTTC, el IGME-CSIC, y los responsables del proyecto SARAI no asumen responsabilidades por cualquier error u omisión contenido en ellos.

### **2.4.4. Calidad y organización de los datos** [Are data quality assurance processes described?]

Para asegurar la calidad de los datos se estructurarán en diferentes carpetas, cuyos nombres identifiquen su contenido y versión. La versión será asignada por un número, decimal o entero, precedida de la letra “v”. Este número se irá incrementando a medida que las carpetas y los ficheros de datos se actualicen. Los datos recopilados tendrán que ser guardados en carpetas que no se modifiquen para asegurar su calidad. Los datos generados serán validados en las zonas piloto con trabajos insitu, y en las zonas donde se dispone de resultados de proyectos anteriores, comparando estos resultados con los obtenidos en el proyecto SARAI. Así mismo, las herramientas matemáticas a las que se hace referencia en el apartado 1.2., donde se describen la relación de los datos con los objetivos del proyecto, permitirán realizar estimaciones de incertidumbre y errores de la mayoría de los productos generados en el proyecto SARAI.

**2.4.5. Período de reutilización de los datos** [How long is it intended that the data remains re-usable?]

El período de tiempo durante el cual los datos albergados en el repositorio Zenodo o en las bases institucionales o servicios de mapas del IGME-CSIC permanecerán reutilizables es indefinido. Los datos con acceso restringido estarán sujetos a las condiciones de los acuerdos de compromiso adquiridos.

**2.4.6. Propiedad intelectual**

La propiedad de los resultados protegibles o no que se generen como consecuencia de la ejecución y desarrollo del proyecto SARAI corresponderá al CTTC y/o el IGME-CSIC, dependiendo de qué parte o en qué proporción se hayan obtenido. En la medida que estos resultados sean susceptibles de protección legal, la titularidad de los Derechos de Propiedad Industrial relativos a las invenciones u otros títulos de propiedad que pudieran derivarse de los citados resultados corresponderá(n) a la(s) institucione(s) que los hayan obtenido, teniendo preferencia para solicitar los correspondientes títulos de protección, apareciendo como inventores o autores aquellos investigadores que hayan contribuido intelectualmente a la obtención de estos resultados. En el caso de que ambas instituciones hayan generado resultados de forma conjunta, ambas serán cotitulares de dichos resultados.

Los datos están protegidos por los correspondientes derechos de propiedad intelectual e industrial y pertenecen al CTTC y/o IGME-CSIC, debiéndose respetar las indicaciones de cita correspondiente. Esta cita deberá realizarse de una de las siguientes maneras: “Origen de los datos: “©Instituto Geológico y Minero de España (IGME-CSIC)”, “Origen de los datos: “©Centre Tecnològic Telecomunicacions Catalunya (CTTC)”, o ambos organismos. No obstante, y de acuerdo con condiciones generales para la reutilización de acuerdo con el ANEXO del RD 1495/2011, de 24 de octubre de desarrollo de la Ley 37/2007 sobre reutilización de la información del sector público, se permite el uso, reproducción total o parcial, distribución, comunicación pública, reutilización o cualquier otra actividad similar o análoga de los datos. La transformación y modificación de la información en su reutilización no podrá suponer ninguna clase de alteración ni desnaturalización de su contenido. Determinadas bases o conjuntos de datos podrán tener unas condiciones específicas de uso, que limiten el acceso y descarga de la información en algunos casos y que estarán recogidos en las condiciones específicas de su licencia de reutilización.

### 3. Asignación de recursos [Allocation of resources]

#### 3.1. Recursos humanos y económicos para garantizar que los datos de investigación cumplan con los principios FAIR [What are the costs for making data FAIR in your project?]

El personal designado para asegurar la correcta gestión de los datos serán los responsables de cada uno de los objetivos del proyecto SARAI, los cuatro IPs que podrán delegar la construcción y almacenamiento de los datos en otro personal del equipo de investigación y del equipo de trabajo, según el cronograma designado en la memoria del proyecto. Además, será necesario formar a personal implicado en el objetivo O4 Generación de una base de datos de deformación InSAR, sobre base datos de series temporales tipo TimescaleDB, tipo SQL que sea escalable para datos de series temporales, que sean multidimensionales.

El personal asignado para el control de la estructura de los datos y metadatos es personal experto en gestión de los datos con formato estándar INSPIRE (área de Sistemas de información del IGME-CSIC).

No existe un presupuesto directo asignado para la gestión de los datos, pero sí de forma indirecta al haber dedicación de expertos técnicos destinada exclusivamente a controlar que los datos de investigación cumplan los principios FAIR. También se han asignado 15.000,0 € para el desarrollo de una web que pueda ser portal de acceso a todos los productos del proyecto SARAI, así como un visor web de los datos. Existe un presupuesto de hardware para el almacenamiento de los datos, consistente en:

- 1 estación de trabajo de sobremesa (6.000,0€)
- 2 portátiles (3.000,00 €)
- Discos duros externos (1.000 €)
- Servidor de almacenamiento de datos (4.000 €)
- Discos duros externos (8.000 €)
- 3 servidores con altas capacidades (12.000 €).

#### 3.2. Responsables de la gestión de datos en el proyecto [Who will be responsible for data management in your project?]

Las personas designadas para la gestión de los datos dependiendo de los objetivos del proyecto son:

- O1 y O2: Oriol Monserrat y Michele Crosetto.
- O4 y O6: Marta Béjar
- O5 y O7: Carolina Guardiola

Estos responsables asegurarán una correcta obtención de los metadatos, la fiabilidad de los datos, su seguridad, así como una correcta colaboración entre las actividades relacionadas con los distintos objetivos.

#### 3.3. Preservación a largo plazo [Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?]

No existe presupuesto explícito en el proyecto SARAI destinado a la preservación a largo plazo de los datos generados, más allá del hardware de almacenamiento descrito en el apartado 3.1. Dentro de las tareas y presupuestos del IGME-CSIC está la

preservación de las bases de datos institucionales. Se evaluará durante el presente proyecto la viabilidad de que estos datos puedan formar parte de dichas bases de datos para garantizar su preservación a largo plazo. Si el resultado de esta evaluación es positivo se solicitará mediante el procedimiento establecido que los datos formen parte de una base de datos institucional.

#### 4. Almacenamiento y salvaguardia de seguridad [Data security]

##### 4.1. Recuperación datos, almacenamiento seguro [What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?]

La recuperación de los datos se garantizará mediante la realización de copias de seguridad periódicas, con preferencia diaria, en los discos duros externos y el servidor de almacenamiento de datos ubicados en el IGME y el CTTC. El almacenamiento de forma segura y la conservación de los datos a largo plazo se garantizarán depositando los datos en los repositorios mencionados en el apartado 2.2.2. La persona responsable de la salvaguarda y restauración de los datos será el responsable de su gestión, o una persona que este responsable delegue.

La base de datos y el catálogo de metadatos del IGME-CSIC se almacenan en servidores que tienen establecidos plan de mantenimiento y seguridad, por lo tanto se adherirán a dicho plan.

La gestión y control del acceso a colaboradores para trabajar con los datos o las transferencias de datos manteniendo la seguridad, integridad y confidencialidad (si es el caso) será gestionada y consensuada por los responsables del proyecto.

#### 5. Aspectos éticos [Ethical aspects]

##### 5.1. Aspectos éticos y protección de datos [Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA). Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?]

Los aspectos éticos del uso de la inteligencia artificial y los datos satelitales de observación de la Tierra son hoy en día un problema abierto. En general el proyecto SARAI garantizará que el uso de los datos dentro de las actividades propuestas esté sujeto a los estándares internacionales científicos de conducta ética (ver referencias en <https://www.csic.es/es/el-csic/etica/etica-en-la-investigacion>). Además, el proyecto SARAI velará por que la aplicación de la inteligencia artificial según la Guía de ética en inteligencia artificial publicada por la unión europea en abril de 2019 (<https://data.europa.eu/doi/10.2759/177365>). Más información sobre la ética de la aplicación de la inteligencia artificial a los datos satelitales puede consultarse en la web del proyecto "Artificial Intelligence for Earth Observation" <https://ai4eo.de/>.

De conformidad con la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD), los datos personales de los miembros del proyecto, sus colaboradores, usuarios de los datos o de las entidades promotoras observadoras serán incorporados a las listas de contactos de los responsables del proyecto, para la correcta gestión del proyecto, la difusión de los resultados y futuras colaboraciones científicas. Se podrá ejercer el derecho de acceso, rectificación, cancelación y

oposición previstos en dicha Ley, dirigiendo un correo electrónico a la dirección de email de los responsables del proyecto. El almacenamiento de dichos datos deberá garantizar la seguridad de dicha información y en ningún caso estos datos serán cedidos a terceros.

## 6. Otros [Others]

### 6.1. Procedimientos, políticas o referencias

El presente PGD se ha desarrollado a partir de la plantilla elaborado por la Universidad de Málaga, que a su vez utilizó como base la herramienta “[DMPonline](https://dmponline.dcc.ac.uk/plans)” (<https://dmponline.dcc.ac.uk/plans> Horizon 2020 DMP).